DOCUMENT RESUME

ED 024 979 CG 003 146

By- Norman, Warren T.

To See Ourself as Others See Us!

Michigan Univ., Ann Arbor.
Spons Agency- National Inst. of Mental Health (DHEW), Bethesda, Md.
Pub Date [68]
Note- 26p.
EDRS Price MF-$0.25 HC-$1.40
Descriptors- *Factor Structure, *Forced Choice Technique, Measurement Techniques, *Peer Acceptance, *Personality Assessment, Self Concept, *Self Evaluation, Sociometric Techniques

A series of studies is reported which explores methodological and procedural variables in interpersonal perception research, and proposes alternative methods to improve personality assessment in this area. Initial studies attempted to corroborate a five-factor model of peer-ratings, and validate self-report ratings on these factors against peer-rating criteria. In the latter studies, a predicted peer-rating was added to the design for comparison with actual peer-rating and self-rating. Discussion follows two paths, evaluation of statistical procedures, and presentation of research results. No conclusions were drawn from these studies, because there was little evidence of convergent and discriminant validity in comparing peer-ratings with self-ratings. With a new sample of acquainted Peace Corps trainees, moderate convergent and discriminant validities were obtained. In a similar sample of unacquainted subjects, the relationships between peer-ratings and self-ratings were lower, and relationships between predicted peer-ratings and both other ratings were lower and less uniform as well. The above findings are viewed as tenuous in light of methodological difficulties. (BP)

OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

"To see ourself as others see us!"

Warren T. Norman[1]

University of Michigan

The research results I will present in this paper--and try to interpret--were drawn from a series of studies carried out over the past eight or nine years. The primary focus of these investigations has not been on the processes and phenomena of interpersonal perception as such. Rather, the emphasis has been on the development of alternative methods to improve the assessment of personality attributes within the normal range of human variation. Indeed, the earlier studies in this series took the results of peer-rating assessments as criteria against which other procedures, stimulus materials, and scoring methods were to be "validated."

Previous findings by Fiske (1949), Borgatta (1960, 1964), Tupes (1957), and Tupes & Christal (1958, 1961) had indicated that a stable factorial structure existed for the given set of peer-rating scales across diverse conditions and populations and that scores derived from these scales predicted to military performance criteria in the Air Force. Subsequent studies have tended to corroborate and extend these findings both as regards the stability of the factorial structure and the predictive power of the scales against a variety of performance criteria in various contexts. The tentative names assigned to the five factors in this structure are: I. Extroversion or Surgency; II. Agreeableness; III. Conscientiousness; IV. Emotional Stability; and V. Culture.

The initial studies in the present series attempted two things; first, to replicate and extend to yet another population the previous findings regarding the factorial structure of the peer-rating scales and second, to develop and validate self-report measures of these dimensions against the peer-ratings considered as (intermediate) criteria. Accounts of these efforts have been published previously (Norman, 1963a and Norman, 1963c) as have the results of some additional and unsuccessful efforts to employ maximal performance tests and preference responses to content irrelevant stimuli to assess these factors (Norman, 1963b). In brief, moderate convergent and discriminant validities were achieved for a five-factor structure by use of content relevant stimuli presented in forced-choice formats. In addition, specially designed scales for detecting faking and a novel scoring procedure for the content factors yielded substantial control over tendencies to distort self-reports on these instruments.

While a moderate degree of success was achieved by these efforts, the convergent validities were neither so high as one might wish for applied purposes nor so large as the respective reliabilities would theoretically permit. Clearly there were systematic (non-random) sources of variance present in the various sets of measure variables

ED024979
CG 003 146

---

[1]The author is grateful to Eva F. Bradford for her assistance in the collection and analysis of much of the data included in this report. The studies reported were supported in part by Research Grant MH-07195 from the National Institute of Mental Health, United States Public Service.

that were unrelated to the personality constructs these measures were intended to assess. Moreover, some of these extraneous determinants of score variance could easily be identified as purely methodological in character.

For example, the peer-ratings were obtained by forcing distributions of nominations by each rater. A consequence of this forced nomination procedure was that all groups of ratees necessarily obtained identical mean scores on all rating scales. Since the self-report measures were not similarly constrained by the procedures used to collect and score them, a systematic difference in method variance components existed and thus served to attenuate validities.

In like manner the (quasi-)counterbalanced, forced-choice format employed for the self-report instruments resulted in mean profiles for all respondents that had (nearly) constant "elevations" or mean levels across scales. No such restraint was present in the peer-rating measures and this presumably limited even further the convergent validities attainable.

Efforts by Norman and Harshbarger (1965) to identify and estimate the magnitudes of the various methodological and substantive constituents of these measure variables and to devise adjustment formulas to bring them into closer mutual alignment met with only modest success. For four of the five personality factors, varying amounts of increase in the correlations were obtained, averaging about .10 points, while for the fifth factor little if any increase resulted from any of the adjustments.

Thus, while it was possible to identify a number of format and procedural components in each of the respective sets of assessments, adjustments or corrections devised to remove these methodological artifacts were only partially successful. There remained substantial portions of reliable variance in each set of assessments still unaligned with that in the other set. Accordingly our attention turned to other potential sources of this residual variation and we were led to consider other means of identifying and estimating the magnitudes of such components.

Since the self-report inventories we had been using yielded scores that were both (quasi-)ipsatized and subject to various amounts of multiple keying of single response alternatives we constructed two new sets of self-report measures that would be free of these artifacts. The first was a set of statistically and conceptually homogeneous keys for an inventory composed of ordinary self-report items. The second set of variables was constructed by the simple expedient of attaching to each of the bipolar peer-rating forms a graphic scale on which the respondent could register a self-rating as well as his estimate of where his peers would place him on this scale.

These measures, together with the peer-rating scales, were administered to several samples of subjects. Small pilot samples of firemen and fraternity seniors, a larger group of fraternity men, a

sample of Peace Corps Trainees and a large group of university student volunteers were tested with various parts of the battery. The major analyses were carried out on the group of 169 fraternity men who responded on all three sets of measures. Only the scales for the first four factors were used in these analyses because of the limitations of the self-report item pool relative to the fifth factor.

The analysis of these data proceeded in two related directions. The more direct approach was based on the rationale that if the subsets of variables within each of the three types did in fact tap the dimensions they had been designed to measure, then a joint analysis of all the variables should yield a factor pattern in which four content factors should emerge, each with three subsets of salients (corresponding to the variables of each type that had been designed to tap that factor). If, in addition, all variables of a given type differed from those of other types in some systematic manner, then one might expect one or more methods factors also to appear in the pattern. Table 1 presents a schematic representation of such a possible factor pattern in which there are four "pure," heteromethod, content factors and a pair of bipolar (differential) method factors. The method factors portrayed contrast peer vs. self sources of data and stimulus presentation and/or response format differences within those measures based on the two types of self-reports. Other kinds and combinations of method factors would also be possible, of course, and those presented in Table 1 are meant only to be illustrative.

There exists a major technical reason for questioning the feasibility of this approach, however. Note that the hypothesized factor pattern given in Table 1 does not conform very well to the simple structure prototype. In particular, note that <u>every</u> variable is presumed to have a complexity greater than one and that the hyperplane counts for the method factors would presumably be very low--in fact, approaching zero in some cases. Such features of a data set are anathema to rotational methods based on simple structure criteria. Since most of the commonly used analytic methods (including Varimax and Biquartimin) are based on such criteria, efforts to rotate data matrices with structures like that given in Table 1 by means of such methods might well fail to attain such a solution.

An attempt was made, however, to determine whether convergent loading patterns could be obtained directly by an application of these standard rotational criteria to the principal axes of each of the monomethod sets of variables as well as those for each heteromethod combination of the variables. The general findings of these analyses were that (a) Varimax and Biquartimin solutions for single (monomethod) sets of variables displayed the expected dimensionality and good approximations to simple structure patterns in all three cases but, (b) neither Varimax nor Biquartimin solutions for joint (heteromethod) combinations of variables gave adequate approximations to the type of pattern presented in Table 1. Only rudimentary indications of the convergent pattern were detectable in some columns of each of these solutions.

Table 1

A Hypothetical Factor Pattern from a Joint Analysis of all of the Variables

		Content Factors				Method Factors	
		I	II	III	IV	M_1	M_2
Peer Rating Variables	1	+				+	
	2	+				+	
	3	+				+	
	4	+				+	
	5		+			+	
	6		+			+	
	7		+			+	
	8		+			+	
	9			+		+	
	10			+		+	
	11			+		+	
	12			+		+	
	13				+	+	
	14				+	+	
	15				+	+	
	16				+	+	
Self Rating Variables	1	+				-	+
	2	+				-	+
	3	+				-	+
	4	+	+			-	+
	5		+			-	+
	6		+			-	+
	7		+			-	+
	8		+			-	+
	9			+		-	+
	10			+		-	+
	11			+		-	+
	12			+		-	+
	13				+	-	+
	14				+	-	+
	15				+	-	+
	16				+	-	+
Questionnaire Variables	1		+			-	-
	2		+			-	-
	3		+			-	-
	4		+			-	-
	5		+			-	-
	6		+			-	-
	7			+		-	-
	8			+		-	-
	9			+		-	-
	10			+		-	-
	11			+		-	-
	12			+		-	-
	13				+	-	-
	14				+	-	-
	15				+	-	-
	16				+	-	-
	17				+	-	-
	18				+	-	-
	19					+	-
	20					+	-
	21					+	-
	22					+	-
	23					+	-
	24					+	-

Note: + and - indicate loadings of salient magnitude.

Table 2 presents the normalized Varimax solutions for the three sets of variables analyzed separately. Table 3 presents the normalized Varimax solution for all three sets of variables analyzed jointly. The similarity of the results in Table 3 with the projected pattern presented in Table 1 is not very great--even if only those columns which most closely approximate the content factors of the hypothetical pattern are considered. Clearly a rotational method based on simple structure criteria (i.e., Varimax) was badly suited to the analysis of these non-simple structure data. Biquartimin solutions which were obtained for these same data were hardly any better.

Fortunately, however, there does exist an alternative rotational method that does permit the investigator to test his prior hypotheses concerning the structure of complex sets of variables such as these. It carries the unfortunate and, as I hope to show, somewhat misleading name of "Procrustes." In this approach the investigator is allowed (and required) to specify fully his hypothesis in the form of a numerical factor pattern. The principal axis solution (or some other derived configuration) is then rotated into the closest possible approximation to this pattern that the intrinsic structure of the data permits. The solution is thus a "fitted" one. Accordingly the danger exists (as the name "Procrustes" suggests) that the data may be "cut" or "stretched" to fit the specified "Procrustean bed," willy-nilly. Horn (1965) has in fact shown that such a possibility exists with "data" that are <u>without any intrinsic structure</u>--i.e., entirely random. However, as the right sides of Tables 4, 5, and 6 show, there are clearly limits to the extent that <u>intrinsically structured data</u> such as those presently under consideration can be wrenched to fit a specified pattern. These "worst" solutions were obtained by specifying a pattern which was chosen to be as thoroughly contrary as possible to that hypothesized to exist in these sets of variables. Clearly these data could not be made to fit, with any detectable comfort, so harsh a bed!

The solutions given on the left of Tables 4, 5, and 6 are those obtained by specifying a "Best" a priori pattern based on the design of these batteries. These oblique solutions are somewhat of an improvement over the corresponding varimax solutions presented in Table 2 although the correlations among the primaries are not very high in most instances.

Thus it was found that by specifying a target pattern to be fitted in each case by a Procrustes procedure very close fits to the a priori specifications could be obtained. What is more, efforts to fit these data to alternative specifications that were markedly disparate from those based on the psychometric design of the variables produced extremely poor fits. Thus, the sometimes expressed opinion that a Procrustes procedure is capable of wrenching just any data matrix to conform to a given specification or target pattern is clearly not a valid one, at least for data as highly structured as these.

Table 2

Normal Varimax Solutions Based on 169 Fraternity Men

a) 16 Peer Rating Variables

Variable No.	Factor I	II	III	IV	h^2
1	86				78
2	72				53
3	74			31	73
4	78		-38		75
5		82			76
6		71		46	72
7		80			77
8		71			74
9	-34		71		71
10			85		83
11		43	72		78
12			87		82
13		49		59	59
14				81	71
15				77	71
16				62	46
	2.85	3.02	2.95	2.59	11.40

b) 16 Self Rating Variables

Variable No.	Factor I	II	III	IV	h^2
1	66				45
2	56				32
3	54				45
4	53				32
5		49			29
6		40		40	32
7		59			40
8	37	52			47
9			58		38
10			79		69
11		31	48		37
12			65		50
13				62	42
14				68	52
15				45	25
16					10
	1.59	1.38	1.84	1.45	6.26

c) 24 Questionnaire Variables

Variable No.	Factor I	II	III	IV	h^2
1	56				41
2	54	-31	33		50
3	65				50
4	51			47	56
5	62			37	53
6	39			47	38
7	43	32			31
8		76			58
9		65			47
10		71			54
11		68			49
12		67			47
13			74		67
14			69		49
15			70		55
16			50		29
17			49	35	42
18			70		57
19			31	68	59
20				71	57
21				83	70
22				79	64
23				64	50
24				58	38
	2.31	2.82	2.94	4.03	12.09

Note--Loadings < $|\pm .30|$ and decimal points for loadings and communalities omitted.

Table 3

16 Peer Rating Plus 16 Self Rating Plus 24 Questionnaire Variables

169 Fraternity Men - Normal Varimax Solution

Variable Set	Variable No.	Factors I	II?	III	PR?	SR?	QR?	h^2
	1	78						72
	2	54						39
	3	64		-36	31			71
	4	69		-40				68
	5		33		72			64
	6				76			68
	7		34		72			71
Peer	8				74			69
Ratings	9			72				66
	10			72	46			77
	11	-32		60	40			70
	12			69	42			72
	13				73			59
	14				70			56
	15				73			58
	16				60			43
	1	68						48
	2	57						36
	3	50				34		42
	4	48						28
	5		40					28
	6					46		35
	7		31			30		29
	8	42				33		37
Self Ratings	9			61				43
	10			72				61
	11			56				40
	12			58		33		47
	13					53		36
	14					67		61
	15					33		33
	16							10

(Table continued on next page)

Table 3 (continued)

16 Peer Rating Plus 16 Self Rating Plus 24 Questionnaire Variables

169 Fraternity Men - Normal Varimax Solution

Variable Set	Variable No.	Factors I	II?	III	PR?	SR?	QR?	h^2
	1	40	30				34	38
	2	41		38				38
	3	51		30				50
	4	34	32				58	57
	5	46					51	51
	6	42					49	44
	7		38					27
	8		77					61
	9		66					48
	10		63					53
	11		63					48
	12		62					46
Questionnaire	13			64			39	57
	14			61				40
	15			63			30	52
	16			55				32
	17			44			43	40
	18			67			31	56
	19						74	59
	20						71	56
	21						79	66
	22						76	62
	23						68	51
	24						61	39
Factor Contributions		5.03	3.84	6.69	5.16	2.23	5.11	28.06

Note--Loadings < $|\pm.30|$ and decimal points for loadings and communalities omitted.

Table 4

16 Peer Rating Variables - 169 Fraternity Men - Procrustes Solutions

Factor No. and Name	Variable No.	"Best" A priori Solution I	II	III	IV	"Worst" A priori Solution A	B	C	D
I Extroversion-Surgency	1	81*				47*	32		43
	2	71*				45	66*	59	83
	3	64*				55	72	46*	40
	4	68*				49	44	82	34*
II Agreeableness	5		70*			50	40		42*
	6		54*				35	38*	60
	7		67*				(26)*		
	8		56*			(20)*		60	32
III Dependability	9			67*		39	(22)*	44	41
	10			76*		(07)*	30	53	58
	11		32	60*		57	68	45	(29)*
	12			80*			31	64*	49
IV Emotional Stability	13				42*			(18)*	31
	14				70*		52	46	81*
	15				68*	(29)*	56		38
	16				51*		(23)*	75	30

Primary Factor Correlations	I				A			
	II	08			B	-86		
	III	-34	25		C	-51	29	
	IV	17	57	14	D	63	-80	-40

Note--* = a specification of .99 for this loading. All other specifications set = .00. Loadings <|±.30| not included except for those in starred positions in which case they are put in ( ). Decimals omitted.

Table 5

16 Self Rating Variables - 169 Fraternity Men - Procrustes Solutions

Factor No. and Name	Variable No.	"Best" A priori Solution I	II	III	IV	"Worst" A priori Solution A	B	C	D
I Extroversion-Surgency	1	65*				49*			34
	2	55*				48	(20)*		
	3	47*				33		37*	49
	4	53*				44			(23)*
II Agreeableness	5		49*						49*
	6		32*		35			46*	47
	7		56*				49*		47
	8		48*			48*	58		61
III Dependability	9			59*			30*		
	10			75*		56*	58		
	11			43*			50		(18)*
	12			61*		54	46	36*	
IV Emotional Stability	13				56*			64*	37
	14				61*			68	43*
	15				45*	(28)*		50	
	16				(17)*		(16)*		

Primary Factor Correlations	I	II	III	IV
	I			
	II	14		
	III	06	21	
	IV	31	29	06

	A	B	C	D
A				
B	-63			
C	-52	65		
D	23	-66	-68	

Note--* = a specification of .99 for this loading. All other specifications set = .00. loadings <|±.30| not included except for those in starred positions in which case they are put in ( ). Decimals omitted.

Table 6

24 Questionnaire Response Variables - 169 Fraternity Men - Procrustes Solutions

Factor No. and Name	Variable No.	"Best" A priori Solution I	II	III	IV	"Worst" A priori Solution A	B	C	D
I Extroversion-Surgency	1	76*	30			76*		64	
	2	50*				38	(27)*	44	
	3	61*				45		53*	
	4	60*				59		69	34*
	5	57*				53*		71	31
	6	38*			40			48*	
II Agreeableness	7	46	42*			65*		56	47
	8		69*			36			68*
	9		68*			43		32*	69
	10		68*				(07)*		64
	11		64*			31*			66
	12		64*			38		33	71*
III Dependability	13			78*			77*	37	36
	14			70*			62		(21)*
	15			75*		(12)*	71	31	
	16			56*			50	(24)*	
	17			60*	33		72*	47	42
	18			56*	31		67	36	49*
IV Emotional Stability	19	30			30*	34	48*	60	35
	20				44*		54	55*	51
	21				68*	(22)*		61	47
	22				48*	30	30	59	46*
	23			51	51*		74*	56	58
	24				40*		30	50*	39

Primary Factor Correlations	I				A			
	II	-05			B	62		
	III	28	08		C	-75	-59	
	IV	24	38	04	D	-30	-42	-11

Note--* = a specification of .99 for this loading. All other specifications set = .00. Loadings <|±.30| not included except for those in starred positions in which case they are put in ( ). Decimals omitted.

However, our major interest was in the extent to which the full heteromethod battery possessed a structure similar to that indicated in Table 1. The Varimax solution presented in Table 3 gave little assurance that such was the case--but perhaps for reasons of inappropriateness already mentioned. Table 7 presents a Procrustes solution for these data based on a specification like that in Table 1. A very clear pattern of heteromethod convergence is reflected in the first four columns of these results. A pair of "flat" method factors contrasting, respectively, peer ratings with self-reports and self-ratings with questionnaire responses within the self-report domain round out the picture.

Whereas the analyses described above have attempted to display the amount of heteromethod convergence in these data directly in terms of the coefficients of the factor patterns, the second approach to the analysis utilized a more indirect, three-stage procedure. First, the variables of each type were separately factor analyzed and rotated--in this case to simple structure solutions (cf, Tables 2 and Tables 4, 5, and 6). Next, factor score estimates were computed for each subject on each factor of each of the separate analyses. And finally, the factor score estimates from all analyses were intercorrelated. If the several sets of variables each yielded measures of the same factors then the correlations obtained in this manner should display what Campbell and Fiske (1958) have termed a convergent and discriminant validity pattern.

For purposes of this analysis, each set of variables actually yielded three sets of factor score estimates. The first two sets were multiple regression estimates based, respectively, on a normalized Varimax solution and a Procrustes rotation of the principal axes. The third set of estimates was based on a simple unit-weighting of the a priori salients for each factor presumed to be present in each set of data.

It is clear from the separate factor tables already presented that the structures of the several sets of measures are highly similar and that, except for a few exceptions, those variables presumed to tap each of the factors do so. It is also apparent that even the orthogonal (normalized Varimax) solutions given in Table 2 approximate a simple structure form rather closely, and the oblique (Procrustes) patterns in Tables 4, 5, and 6 do so to an even higher degree.

But the proof of convergence lies not in comparisons among the separately obtained factor patterns but rather in the correlations among the sets of factor score estimates. Table 8 presents the multitrait-multimethod matrices based on the three factor score estimation procedures utilized. It also presents the correlations between estimation procedures for each trait-by-data-source combination.

The first thing to note about the results in Table 8 is that the methods of deriving factor score estimates (i.e., multiple regression estimates from Varimax and Procrustes solutions and unit-weighted, a priori salients) all yield very highly similar results. This is

Table 7

16 Peer Rating, 16 Self Rating, and 24 Questionnaire Variables
169 Fraternity Men - Procrustes Solution

Variable Set	Variable No.	Factors I	II	III	IV	M_1	M_2
Peer Ratings	1	74*				(09)*	
	2	52*				(22)*	
	3	54*				(23)*	
	4	59*				(-03)*	
	5		50*			37*	
	6		43*			41*	
	7		47*			40*	
	8		48*			(29)*	
	9			67*		(-15)*	
	10			71*		(17)*	
	11			57*		(20)*	
	12			67*		(24)*	
	13		34		33*	34*	
	14				45*	36*	
	15				36*	48*	
	16				33*	35*	
Self Ratings	1	63*				(-16)§	(08)∝
	2	55*				(-14)§	(01)∝
	3	37*				(-24)§	(28)∝
	4	42*				(-26)§	(13)∝
	5		48*			(-22)§	(11)∝
	6		42*			(-21)§	(27)∝
	7		41*			(-19)§	(25)∝
	8	30	(28)*			(-20)§	35∝
	9			58*		(-17)§	(-08)∝
	10			74*		-33§	(27)∝
	11			54*		(-18)§	(06)∝
	12			60*		-30§	(21)∝
	13				40*	-34§	(24)∝
	14				63*	-30§	30∝
	15				41*	(-04)§	(23)∝
	16				(16)*	(-16)§	(18)∝

(Table continued on next page)

Table 7 (continued)

16 Peer Rating, 16 Self Rating, and 24 Questionnaire Variables
169 Fraternity Men - Procrustes Solution

Variable Set	Variable No.	Factors I	II	III	IV	M_1	M_2
Questionnaire	1	44*				-36§	(-23)§
	2	41*		39		(-01)§	(-00)§
	3	59*				(01)§	(-26)§
	4	41*				-40§	-36§
	5	36*				(-14)§	-34§
	6	49*				(-15)§	-38§
	7		38*			-32§	(-08)§
	8		70*			(-22)§	(-27)§
	9		63*			(-19)§	(-21)§
	10		61*			(-23)§	(-11)§
	11		59*			(-25)§	(-18)§
	12		57*			(-29)§	(-20)§
	13			61*		(-22)§	(-17)§
	14			61*		(-19)§	(16)§
	15			61*		(-29)§	(-03)§
	16			54*		(-09)§	(01)§
	17			39*		(-19)§	(-28)§
	18			63*		(-27)§	(-16)§
	19				44*	-30§	-39§
	20				44*	(-29)§	-38§
	21				59*	(-24)§	-42§
	22				56*	(-27)§	-39§
	23				41*	(-19)§	-41§
	24				42*	(-13)§	-35§

Primary Factor Correlations	I						
	II	-11					
	III	-04	04				
	IV	28	23	08			
	M_1	06	29	12	16		
	M_2	22	-03	-21	-04	13	

Note--* = a specification of .70 for this loading; ∝ = a specification of .50 for this loading; § = a specification of -.50 for this loading. All other specifications set = .00. Loadings < |±.30| not included except for those in non-zero specification positions, in which case they are put in ( ). Decimals omitted. M_1 contrasts Peer vs. Self sources of data. M_2 contrasts Rating vs. Questionnaire formats within the Self source.

Table 8

Correlations Among Factor Score Estimates from Three Data Sources and Three Methods of Derivation

N = 169 Fraternity Males

Varimax

			Peer Ratings				Self Ratings				Questionnaire Scales			
			I	II	III	IV	I	II	III	IV	I	II	III	IV
Varimax	PR	I	(94)	All zero										
		II	01	(93)										
		III	-05	04	(95)									
		IV	03	10	01	(91)								
	SR	I	[56]	-15	-12	02	(83)	All zero						
		II	05	[29]	06	05	06	(80)						
		III	-07	-05	[63]	-22	-00	09	(89)					
		IV	11	06	-04	[30]	10	12	-01	(82)				
	QR	I	[39]	-10	14	03	[53]	07	01	03	(89)	All zero		
		II	-23	[44]	01	01	-13	[34]	-04	10	-01	(92)		
		III	-14	-01	[55]	-10	-12	00	[70]	01	05	-01	(92)	
		IV	10	-04	11	[06]	-01	-08	08	[34]	07	02	04	(94)
Procrustes		I	98				57				35			
		II		94				30				39		
		III			97				60				54	
		IV				95				30				07
		I	56				98				52			
		II		26				98				31		
		III			62				99				69	
		IV				30				98				32
		I	37				45				91			
		II		42				32				99		
		III			57				69				98	
		IV				05				35				97
Unit Weight	PR	I	96				55				33			
		II		94				28				39		
		III			96				60				53	
		IV				94				29				06
	SR	I	51				96				53			
		II		31				92				36		
		III			62				96				69	
		IV				32				89				31
	QR	I	39				47				89			
		II		41				35				97		
		III			57				69				96	
		IV				06				33				97

Table 8 (continued)

Correlations Among Factor Score Estimates from Three Data Sources and Three Methods of Derivation

N = 169 Fraternity Males

			Procrustes											
			Peer Ratings				Self Ratings				Questionnaire Scales			
			I	II	III	IV	I	II	III	IV	I	II	III	IV
Varimax	PR	I												
		II												
		III												
		IV												
	SR	I												
		II												
		III												
		IV												
	QR	I												
		II												
		III												
		IV												
Procrustes	PR	I	(87)	08	-34	17								
		II	09	(77)	25	57								
		III	-38	28	(86)	14								
		IV	19	64	16	(76)								
	SR	I	[57]	07	-23	08	(80)	14	06	31				
		II	11	[28]	09	16	24	(78)	21	29				
		III	-15	01	[59]	-14	04	25	(88)	06				
		IV	21	16	-07	[31]	39	35	-00	(79)				
	QR	I	[31]	-07	18	04	[45]	08	21	18	(79)	-05	28	24
		II	-18	[38]	12	12	-05	[30]	-01	17	00	(87)	08	38
		III	-15	-04	[53]	-05	-03	04	[68]	04	47	-01	(84)	04
		IV	04	08	16	[08]	04	03	12	[33]	43	34	25	(83)
Unit Weight	PR	I	99				56				28			
		II		99				17				38		
		III			99				58				52	
		IV				98				30				08
	SR	I	52				97				44			
		II		34				54				35		
		III			60				97				69	
		IV				31				90				32
	QR	I	34				49				96			
		II		39				33				97		
		III			55				67				98	
		IV				08				31				98

Table 8 (continued-2)

Correlations Among Factor Score Estimates from Three Data Sources and Three Methods of Derivation

N = 169 Fraternity Males

			Unit Weight											
			Peer Ratings				Self Ratings				Questionnaire Scales			
			I	II	III	IV	I	II	III	IV	I	II	III	IV
Varimax	PR	I												
		II												
		III												
		IV												
	SR	I												
		II												
		III												
		IV												
	QR	I												
		II												
		III												
		IV												
Procrustes	PR	I												
		II												
		III												
		IV												
	SR	I												
		II												
		III												
		IV												
	QR	I												
		II												
		III												
		IV												
Unit Weight	PR	I												
		II	08											
		III	-44	30										
		IV	16	60	16									
	SR	I	[51]	-11	-25	01								
		II	09	[32]	06	22	18							
		III	-26	-01	[60]	-14	-01	20						
		IV	12	12	05	[31]	22	35	10					
	QR	I	[31]	-03	11	05	[49]	11	14	21				
		II	-20	[38]	17	15	-01	[38]	10	13	13			
		III	-19	-02	[54]	-05	-05	08	[69]	13	33	11		
		IV	02	02	17	[07]	-01	-00	17	[31]	50	21	34	

Note--Values along main diagonal (in parentheses) for Varimax and Procrustes sets are multiple Rs of factor score estimates against factors. Values in mono-source, mono-rotation triangles ◺ above main diagonal for Varimax and Procrustes sets are the correlations among the primary factors.

reflected most clearly in the underlined diagonals of correlations between derivation methods for each of the trait-by-data source combinations. These 36 values range from .89 to .99 with an overall median of .97. The correlations between the Varimax and the Procrustes estimates and those between the Procrustes and the unit-weighted estimates are slightly higher than those between the Varimax and the unit-weighted scores. Thus it would appear that a choice among these procedures for estimating scores of persons on the factors that are implicit in these several data sources is not a very crucial consideration.

The more fundamental findings, however, are contained with the multitrait-multisource submatrices of Table 8. There one will note that the validity diagonals contain vaues which, with one common exception in each submatrix, portray a moderately high level of convergent and discriminant validity. For example, in the Varimax submatrix, only the validity for Factor IV between the peer-ratings and questionnaire sources fails to exceed all of the relevant heterotrait-monosource and heterotrait-heterosource entries. Indeed, the magnitudes of the validities not only stand out from the (generally low) heterotrait values but they are in several instances quite substantial numerically as such estimates generally go. While there is apparently still some reliable variance in each of these sets of measures that is not common across data sources (as witness the multiple correlations in parentheses along the main diagonal relative to the corresponding validities) there is nonetheless an appreciable proportion of common variance across sources, ranging to near 50% in a few cases. And finally, as a further reflection of the comparatively small proportions of unique variance (specificity plus error) in these factor measurements and those based on the Procrustes method, note the close correspondence between the monosource, factor measurement correlations (just below the main diagonal) and the primary factor correlations (presented just above the main diagonal) as obtained from the various factor analyses and rotations done earlier. The latter reflect only common factor sources of variance while the former are affected by all variance components present in the measure variables.

The results based on the Procrustes solutions are a bit more difficult to summarize easily because of the obliquities among the factors within each monosource analysis. Note, however, that each validity can be compared with twelve heterotrait entries. A ratio of the number of these comparisons in which the validity entry is the larger to the total number (twelve) gives a rough index of the extent to which discriminant validity obtains. If discriminant validity is absent, this index should be about .5 and it should approach unity as discriminant validity increases. With the exception of the correlation between the peer rating and the questionnaire measure of Factor IV where this index is .25, the remaining eleven validities have discriminant indices that range from .67 to 1.00 with a median for all twelve of .83. Discriminant validity is obviously best for Factor III and poorest for Factor IV. Interpretations of the somewhat lower magnitudes of these validities (relative to those

based on the Varimax solutions) should be conditioned in terms of the multiple correlations along the main diagonal which are also somewhat lower for these oblique solutions.

Finally, the results in the unit-weight submatrix correspond very closely to those based on the Procrustes solutions and accordingly warrant little additional comment. It is, however, rather astonishing just how closely the results of this very simple method match those obtained by the very arduous and complex procedures entailed by the multiple regression method based on the Procrustes solutions. Apparently one can do a relatively good job of measuring factor scores of persons very simply if he takes the trouble to construct adequate measure variables and to constitute heterotrait batteries to appropriately map the domain of interest.

The most glaring failure to obtain evidence of convergent and discriminant validity in these analyses occurred for Factor IV, Emotional Stability, as measured by the peer ratings and the questionnaire scales, respectively. Actually each of these measurements showed a positive (albeit, moderate) relationship to the measurement of this factor based on self ratings. But apparently the combined effects of source differences together with stimulus presentation and response format differences, were, in this one area of content, sufficient to wash out content convergence.

Whether this result is simply a consequence of poor measure variable design or whether the psychological processes that bear on assessing this molar attribute are markedly different when the object is the self vs. when it is others is not entirely clear. More than likely it is some combination of the two.

It is interesting to note, however, that a factor similar to Emotional Stability (or its bipolar reflection) has been identified as a major source of variance in many of the most widely used personality inventories (cf. Block's recent analysis of the MMPI, 1965). It is also true that previous efforts to establish convergent validity for the scales of such inventories against external rating criteria have frequently been highly discouraging. Such results have occasionally been interpreted as a generalized characteristic and limitation of personality measurement as a whole. It is encouraging to speculate that these previous failures, including the one reported here, may reflect merely on the difficulty of assessing one (albeit, an important one) out of many personality factors of general interest and importance.

While each group of factor score estimates (derived from either the Varimax or the Procrustes solution) was based on independent sets of responses these data were all obtained from the same subjects. Thus while the validities presented in Table 8 are not "fitted" in any sense that capitalizes on random error variation there is a question concerning how widely these functions can be generalized to other samples from this or other populations. Tentative indications based on the two small pilot samples mentioned earlier are presented in Tables 9 and 10. Some appreciable drops in convergent

validities can be noted; particularly in the case of Factor III for the sample of firemen. On the other hand, the convergent values for Factor IV are actually higher for both of these groups than they were for the original sample on which the factor score weights were derived. Additional samples from more diverse subpopulations will be required, of course, before any very final evaluation concerning the generalizability of these findings can be made.

The results presented thus far have dealt mainly with the relationship between peer-rating assessments and self-reports derived either from graphic rating scales or itemized inventory responses. As mentioned earlier, however, some of the samples of respondents were also asked to estimate where on the graphic scales their peers would rate them. For two of these samples, a group of 73 Peace Corps trainees and a group of 84 previously unacquainted university undergraduates, these predicted peer-ratings were also separately factor analyzed and five-factor solutions were derived by normal varimax rotations. Multiple-regression estimates of factor scores were computed for each subject from each of the three sets of variables separately and these factor-score estimates from each of the three data sources were then intercorrelated for each sample. The results are presented in Table 11 (taken from a previously published paper; Norman & Goldberg, 1966).

The correlations for the Peace Corps trainee sample given at the top of Table 11 display a pattern of moderate convergent and discriminant validity between the peer-ratings and self-ratings similar to that previously obtained with other groups of close associates. The previously unacquainted subjects in the second sample, however, yielded lower values in the convergent validity diagonal with those for Factors II and IV failing even to reach statistical significance.

But perhaps the more interesting findings are those between the predicted peer-ratings and the other two sets of variables. For the Peace Corps trainees these predicted peer-ratings correlate higher with the actual peer ratings in every case than do the corresponding self-ratings. In addition, the predicted peer-ratings and the self-ratings are themselves highly correlated in each case, ranging from .69 for Factor II to .88 for Factor I. While these latter findings are surely partly a function of a common item and response format and a common source (the individual respondents) it should be borne in mind that they are, in each case, also based on only a single response per subject per scale. These data are accordingly less reliable than are the peer-rating measures which are based on averages over multiple respondents on each scale.

Thus it would seem that these respondents expect their peers to rate them slightly differently than they rate themselves. In this expectation they are correct in a manner and direction that reflect a degree of perceptual accuracy. However, they expect their peers' ratings of them to conform much more closely to their own self-ratings than does in fact occur.

Table 9

Validity Generalization of the Factor Score Coefficients to a Sample of 56 Fraternity Seniors

			Varimax								Procrustes							
			Peer Ratings				Questionnaire Scales				Peer Ratings				Questionnaire Scales			
			I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
Varimax	PR	I																
		II	02															
		III	11	-15														
		IV	-35	02	-18													
	QR	I	[31]	09	-00	-17												
		II	-28	[45]	26	14	11											
		III	-25	-11	[19]	11	-08	-10										
		IV	-04	14	-12	[30]	06	11	10									
Procrustes	PR	I	98				31											
		II		96				42			08							
		III			97				24		-27	10						
		IV				93				33	-07	56	-00					
	QR	I	24				90				[24]	06	00	01				
		II		46				99			-19	[43]	-15	25	09			
		III			17				98		-22	-04	[21]	09	39	-07		
		IV				31				97	-04	27	-09	[35]	44	40	29	

Note--Factor score coefficients derived from independent sample of 169 fraternity males. Decimals omitted.

Table 10

Validity Generalization of the Factor Score Coefficients to a Sample of 39 Ann Arbor Firemen

			Varimax								Procrustes							
			Peer Ratings				Questionnaire Scales				Peer Ratings				Questionnaire Scales			
			I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
Varimax	PR	I																
		II	-06															
		III	03	17														
		IV	-43	05	28													
	QR	I	[29]	-07	-11	-09												
		II	-30	[44]	-01	13	-15											
		III	-10	15	[-01]	05	18	18										
		IV	-16	01	04	[17]	-13	27	15									
Procrustes	PR	I	99				31											
		II		95				41			-09							
		III			97				05		-28	50						
		IV				93				16	-27	62	54					
	QR	I	23				94				[25]	-07	-13	-04				
		II		42				99			-27	[40]	09	23	-01			
		III			-03				97		-06	10	[00]	06	58	20		
		IV				19				98	-20	14	09	[20]	23	53	35	

Note--Factor score coefficients derived from an independent sample of 169 Fraternity males. Decimals omitted.

The results from the second sample where the subjects were not previously acquainted with one another are, not surprisingly, much more mixed and irregular in their pattern. Not only are the relationships between peer-ratings and self-ratings lower for these subjects than has been previously observed for samplesof close associates, but their predicted peer ratings are also less highly and less uniformly related to both the actual peer-ratings and self-ratings as well. These subjects knew they were being rated by virtual strangers and the correlations of their predicted peer-ratings with the other sets of measures no doubt reflect this fact.

As you are all aware, the title of this presentation was taken from Robert Burns' ode, "To a Louse" and is the second line of the couplet:

> O wad some Pow'r the giftie gie us
> To see oursels as others see us!

Whether our subjects yearned so deeply as the illustrious Scot for such a gift seems doubtful; our results indicate that at least not all of them were richly blessed. But it would appear that they were not wholly unaware either of how others saw them and, indeed, they agreed to a moderate extent with these external judgments. On the other hand, they clearly projected onto their peers an unwarranted amount of agreement with their own self appraisals. Perhaps this is what Burns' lament was all about.

Reprise and Outlook

At the outset of this research program we were willing to consider the collective perceptions of others as reflected in the peer-ratings as a set of criteria against which self-report measures could be validated. But a little thought and several sets of such data (e.g., Norman and Harshbarger, 1965; and Passini and Norman, 1966) revealed a number of methodological and substantive deficiencies that exist in these measurements, also. Constraints owing to forced nomination formats and procedures, interrater disagreements based on differential or limited exposure to the ratees, and ambiguities of the trait descriptors all serve to attentuate the amount of construct valid variance available in the peer ratings. When one considers in addition the potential distortions which accrue to commonly used analysis methods that, however conventional, are more or less ill-suited for these sorts of data, it is hardly surprising that convergent relationships rise no higher than those thus far obtained.

I don't wish to be recorded as an apologist for these and previous sets of findings. But I do think that we have been making

Table 11

Intercorrelations of Factor-Score Estimates Derived from Peer Ratings, Self-Ratings, and Predicted Peer Ratings for Two of the Samples

	Peace Corps trainees (N = 73)[a]									
	Peer ratings					Self-ratings				
	I	II	III	IV	V	I	II	III	IV	V
Self-ratings										
I	[54]	08	-07	-12	00					
II	01	[27]	13	-05	-11					
III	-00	-13	[47]	-17	-19					
IV	22	-07	-27	[32]	-26					
V	04	-20	-16	-25	[45]					
Predicted peer ratings										
I	[62]	16	-18	-06	11	[88]	-06	-09	02	01
II	03	[41]	17	09	-02	-10	[69]	-13	-04	-14
III	07	-08	[56]	-12	-01	13	19	[81]	-07	16
IV	20	-04	-17	[33]	-26	08	13	00	[83]	04
V	09	-24	-15	-14	[54]	-04	-04	-22	-00	[78]

	Passini & Norman Ss (N = 84)[b]									
	Peer ratings					Self-ratings				
	I	II	III	IV	V	I	II	III	IV	V
Self-ratings										
I	[38]	-22	01	15	-11					
II	-02	[15]	07	-27	-01					
III	01	10	[34]	-09	-01					
IV	-11	-17	05	[02]	02					
V	-09	-06	-10	06	[32]					
Predicted peer ratings										
I	[37]	-13	-10	-13	-00	[50]	09	-13	-15	03
II	-19	[15]	02	-20	-17	-25	[25]	03	34	07
III	04	14	[50]	-13	-05	23	13	[66]	-03	-10
IV	22	-11	-09	[26]	04	50	05	-07	[38]	15
V	-15	-02	06	09	[26]	06	09	-09	01	[38]

[a] For N = 73, $r \geq .195$ is significant at $p < .05$ (1-tailed) and $r \geq |\pm .232|$ is significant at $p < .05$ (2-tailed).

[b] For N = 84, $r \geq .181$ is significant at $p < .05$ (1-tailed) and $r \geq |\pm .215|$ is significant at $p < .05$ (2-tailed).

Decimals omitted.

progress in recent years on the development of personality assessment instruments and methods, on identifying and estimating the magnitudes of (and relations among) various methodological and substantive components of the score variables with which we work, and on bringing a more reasonable and sophisticated attitude to bear on the analysis, interpretation and use of these kinds of data for theory development and applied purposes. But we have neither exhausted the possibilities inherent within our present approaches nor have we fully exercized the ingenuity and perseverance needed to invent new and better methods for assessment and for analyzing these sorts of data.

In this latter regard it seems to me that the day has passed when anything of much value is going to be learned by "one-shot" studies using just any conveniently available inventory, sample and data analysis procedure. The need at the present is for systematic investigations that are extensive in scope, that are both multivariate and multi-method in design, and that span an appreciable range of variation on developmental, demographic, and situational facets.

It is unlikely also that "canned" analysis programs are going to suffice for processing the kinds of data that these more elaborate studies will yield. Such methods were, by and large, designed for other, and simpler, models and they often serve effectively for such uses. But their suitability for sorting out the diverse kinds of information implicit in complex batteries administered to stratified samples under multiple task conditions is, in many cases, highly questionable.

The general point I am trying to make in closing is that there is nothing simple, routine, or automatic about research on personality assessment these days; either in the design of investigations or in the analysis and interpretation of results. If we are to make the progress that now seems possible in this area we will have to have the courage and perseverance to undertake large and complex investigations on a programmatic basis and the ingenuity to adapt or invent methods of analysis to the demands that such research designs present.

Block, J. The challenge of response sets. New York: Appleton-Century-Crofts, 1965.

Borgatta, E. F. The stability of interpersonal judgments in independent situations. Journal of Abnormal and Social Psychology, 1960, 60, 188-194.

Borgatta, E. F. The structure of personality characteristics. Behavioral Science, 1964, 9, 8-17.

Fiske, D. W. Consistency of the factorial structures of personality ratings from different sources. Journal of Abnormal and Social Psychology, 1949, 44, 329-344.

Horn, J. L. On subjectivity in factor analysis. Paper read at the meetings of the Society of Multivariate Experimental Psychology, Madison, Wisc., November, 1965 (ditto).

Norman, W. T. Personality measurement, faking and detection: An assessment method for use in personnel selection. Journal of Applied Psychology, 1963, 47, 225-241. (a)

Norman, W. T. Relative importance of test item content. Journal of Consulting Psychology, 1963, 27, 166-174. (b)

Norman, W. T. Toward an adequate taxonomy of personality attributes: replicated factor structure in peer nomination personality ratings. Journal of Abnormal and Social Psychology, 1963, 66, 574-583. (c)

Norman, W. T. & Goldberg, L. R. Raters, ratees, and randomness in personality structure. Journal of Personality and Social Psychology, 1966, 4, 681-691.

Norman, W. T. & Harshbarger, T. R. Matching components of self-report and peer-nomination personality measures. Psychometrika, 1965, 30, 481-490.

Passini, F. T. & Norman, W. T. A universal conception of personality structure? Journal of Personality and Social Psychology, 1966, 4, 44-49.

Tupes, E. C. Relationships between behavior trait ratings by peers and later officer performance of USAF Officer Candidate School graduates. Lackland Air Force Base, Texas: Air Force Personnel and Training Research Center, October, 1957. (Research Report AFPTRC-TN-57-125, AD-134 257.)

Tupes, E. C. & Christal, R. E. Stability of personality trait rating factors obtained under diverse conditions. Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Center, May, 1958. (Technical Note WADC-TN-58-61, ASTIA Document No. AD-151 041).

References (continued)

Tupes, E. C. & Christal, R. E. Recurrent personality factors based on trait ratings. Lackland Air Force Base, Texas: Personnel Laboratory, Aeronautical Systems Division, May, 1961. (Technical Report ASD-TR-61-97.)